Research Article

An Efficient Normalized Rank Based SVM for Room Level Indoor WiFi Localization with Diverse Devices

Yasmine Rezgui, Ling Pei, Xin Chen, Fei Wen, and Chen Han

Shanghai Key Laboratory of Navigation and Location-Based Services, School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

Correspondence should be addressed to Ling Pei; ling.pei@sjtu.edu.cn

This paper proposes an efficient and effective WiFi fingerprinting-based indoor localization algorithm, which uses the Received Signal Strength Indicator (RSSI) of WiFi signals. In practical harsh indoor environments, RSSI variation and hardware variance can significantly degrade the performance of fingerprinting-based localization methods. To address the problem of hardware variance and signal fluctuation in WiFi fingerprinting-based localization, we propose a novel normalized rank based Support Vector Machine classifier (NR-SVM). Moving from RSSI value based analysis to the normalized rank transformation based analysis, the principal features are prioritized and the dimensionalities of signature vectors are taken into account. The proposed method has been tested using sixteen different devices in a shopping mall with 88 shops. The experimental results demonstrate its robustness with no less than 98.75% correct estimation in 93.75% of the tested cases and 100% correct rate in 56.25% of cases. In the experiments, the new method shows better performance over the KNN, Naïve Bayes, Random Forest, and Neural Network algorithms. Furthermore, we have compared the proposed approach with three popular calibration-free transformation based methods, including difference method (DIFF), Signal Strength Difference (SSD), and the Hyperbolic Location Fingerprinting (HLF) based SVM. The results show that the NR-SVM outperforms these popular methods.

1. Introduction

With the fast growing of ubiquitous computing, the rapid advances in mobile devices, and the availability of wireless communication, wireless indoor positioning systems have become very popular in recent years. Moreover leveraging public infrastructure has many advantages such as cost efficiency, operational practicability, and pervasive availability.

Various short-range radio frequency technologies are broadly researched to address positioning task in GNSS denied area, for instance, Radio Frequency Identification (RFID) [1], Wireless Local Area Network (WLAN, a.k.a. WiFi) [2, 3], Bluetooth [4, 5], ZigBee [6], Ultra Wide Band (UWB) [7], and cellular networks [8]. All these means are high potential alternatives to indoor positioning. Meanwhile, indoor localization based on signals of opportunity (SoOP) is still a challenging task, since it requires stable interior wireless signals and adequate adaptation of the signals which are not originally designed for positioning purpose.

Fingerprinting method [9] is one of the most popular and promising indoor positioning mechanisms. It is a technique based upon existing WiFi infrastructure and thus requires no dedicated infrastructure to be installed. It allows positioning by making use of signal characteristics using the signature matching technique. Fingerprinting technique is accomplished by two steps. Firstly, it stores WiFi signatures from different radio wave transmitters for each reference position. Then, it compares the current signature of a device with prerecorded signatures to find the closest match.

Different techniques and solutions have been proposed providing a varying mix of resolution, accuracy, stability, and challenges. Early examples of a positioning system that uses fingerprinting are RADAR [2] and Horus [10] systems. Horus system is based on the probabilistic approach which considers the statistical characteristics of the RSSIs and their distribution. The Lognormal distribution [10], Weibull function [5], Gaussian distribution [11], and the double peak

Gaussian distribution [12] were used to model the RSSI distribution. However, RADAR system was the first introducing the fingerprinting technique based on the deterministic approach [13] using *K*-Nearest Neighbor method (KNN) [14]. Nowadays, the use of machine learning algorithms (ML) as in [15] has increasingly gained more popularity in indoor navigation domain because of their witnessed robustness; among them are Naïve Bayes classifier [16, 17], Support Vector Machine (SVM) [17, 18], Random Forest [18, 19], and Neural Network [20, 21]. However, their model generalization to different user's terminals has seldom been considered.

RSSI fluctuation and diversiform smartphones can significantly degrade the positioning accuracy of WiFi localization systems, as well as the patterns between training and testing signature vectors. The WiFi device used to train the radio map during the calibration phase may especially differ from the ones used during the positioning phase. WiFi modules from different providers have varying receive signal gains which make the RSSI vary using different devices at the same location [22]. This hardware variance problem is not only limited to differences in the WiFi chipsets used by training and tracking devices. Besides, it arises when the same WiFi chipsets are connected to different antenna types and/or packaged in different encapsulation materials.

To address this issue, several studies have proposed methods to improve the robustness of positioning systems against mobile devices heterogeneity. For example, the use of an unsupervised learning [23], the Hyperbolic Location Fingerprinting (HLF) [24], the DIFF method [25], and the Signal Strength Difference method (SSD) [26] are the representative ones. In this paper, a new method applying an intermediate step of absolute RSSI value transformation by making use of ML algorithms is proposed.

The indoor navigation market addresses various applications like Location Based Services (LBS), the guidance of firefighters, and people management. Moreover, it can benefit from the room level accuracy [27–29] to extract statistical information which can be deployed to better market to customers, find hangouts in airports, and even most popular shop in shopping malls, and so forth. In this work, we focus on a room localization which is the prediction of an occupied room in which the mobile device is currently in. This task is more practical to attain since it does not require the floor map of the desired indoor environment which is not always available in practice.

To achieve a high room level accuracy based WiFi fingerprinting technique, we have investigated various database definitions and the impact of signature clustering in our previous work [30], with respect to relevant requirement parameters. In this paper, we propose a normalized rank transformation based SVM approach to solve the issue of mobile terminal diversity during the positioning phase. We consider and compare between the performances of the aforementioned ML algorithms and calibration-free transformations to validate our solution. This approach aims at generalizing the use of the preconstructed radio map derived from one device to manifold devices and guarantee localization accuracy in the meantime.

2. Normalized Rank Transformation

Aiming to mitigate the effect of signal fluctuation and hardware variance issue, we introduce the intermediate normalized rank transformation step to freeze the variation of the RSSI. This solution has been defined principally to deal with SVM classifier. Moving from RSSI value based analysis to the normalized rank transformation based analysis, the principal features are prioritized and the dimensionalities of signature vectors are considered.

2.1. Rank Transformation. In the conventional classification based on SVM or any other ML algorithms for fingerprinting-based indoor localization, the features are defined as the signal strength received from all the visible access points (APs) to build the model. However, the received signal strength is inherently time varying at a specific location. Moreover, device diversity will impact both learning and positioning phases.

In order to find the perfect match between the user location and the predefined locations in the radio map using SVM classifier, as well as attenuate the susceptibility of decision boundaries to RSSI variation, the enhancement of data learning to build a robust model is a key. The improvement of the accuracy is related to the strong decision boundaries between the different classes. The input variables are redefined and the absolute RSSI values are replaced first by their corresponding rank values. Let $P_{1:M}$ be the RSSI vector of the M visible APs and S_i the corresponding absolute RSSI value of the AP(i) such that

$$P_{1:M} = \{S_1, S_2, \ldots, S_M\} \tag{1}$$

$$P'_{1:M} = \{S'_1, S'_2, \ldots, S'_M\} \tag{2}$$

$$S'_1 \leq S'_2 \leq \cdots \leq S'_M \tag{3}$$

$$\Psi_i = \begin{cases} \Psi_{i-1} + 1 & \text{for } S'_{i-1} \neq S'_i \\ \Psi_{i-1} & \text{otherwise.} \end{cases} \tag{4}$$

$P'_{1:M}$ is the new defined RSSI vector by rearranging the APs. The rearrangement is based on (3), in which the throughputs of received signals are in ascending order. Equation (4) explains how to decide the rank vector. If RSSIs values are distinct ($S'_{i-1} \neq S'_i$, i indicates the position of the AP in the M dimensional vector), successive numbers denoted by Ψ_i will be assigned. Otherwise, the same rank will be allocated. The initial value is equal to one and Ψ_M does not have to be equal to the dimension of the initial RSSI vector. By adopting this procedure, the most reliable APs with high RSSI are tagged with the high rank values.

2.2. Normalized Rank Transformation. The observed set of APs is not fixed over time at every calibration point and, consequently, the dimension of the transformed rank vectors.
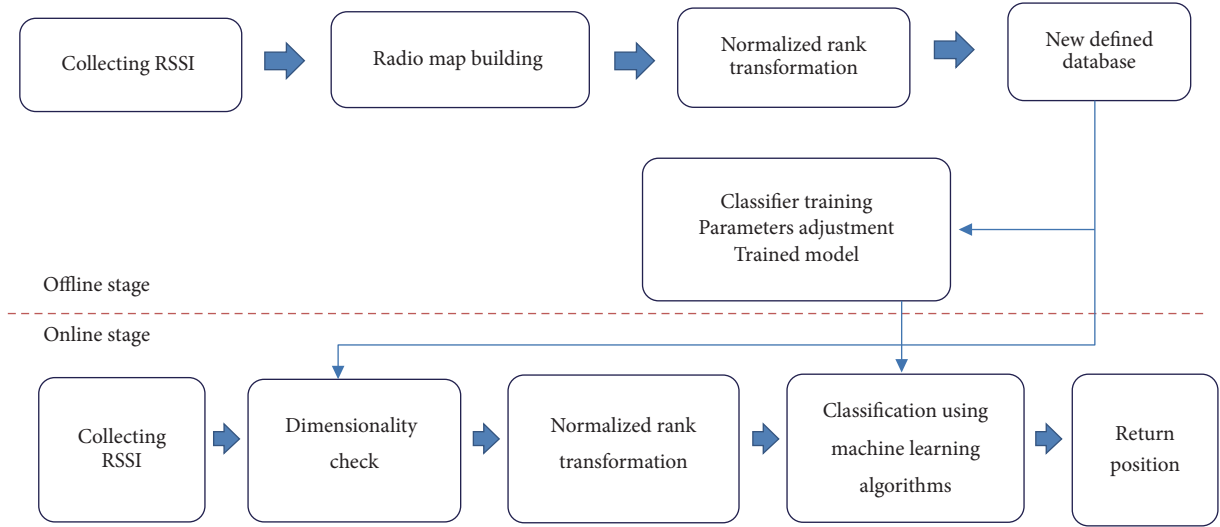
Figure 1: Proposed system architecture.

Hence, the new formulated rank vector has to be adapted to the varying assigned tags and needs to be normalized.

$$NR = \{\lambda\Psi_1, \lambda\Psi_2, \ldots, \lambda\Psi_M\} \quad \text{such that } \lambda = \frac{1}{\Psi_M} \quad (5)$$

$$NR_i \in [0, 1] \quad \text{such that } NR_i = \lambda\Psi_i. \quad (6)$$

Equation (5) defines the normalized rank vector which is denoted as NR and satisfies (6), where λ is the normalization factor corresponding to the inverse of the highest assigned rank Ψ_M in the transformed rank vector. This normalization step will result in keeping the prioritization of the most reliable APs. Furthermore, it helps attenuating the effect of the noisy points as the hyperplanes are always dominated by the high values.

3. Proposed Method

In this section, we give more details about the proposed approach which is composed of two main stages: the offline stage and the online stage, as illustrated in Figure 1.

3.1. The Offline Stage. This stage consists of 3 major steps: *Step 1:* data collection and association; *Step 2:* normalized rank transformation; *Step 3:* model building and parameters adjustment.

Step 1 (data collection and association). Selection of the sampling Reference Point within the region of interest is required to further collect the RSSI by a mobile node from all the available APs. Since the propagation of the radio signal in indoor environments is very complicated, several observations in the same location are needful. A single signature for each location is assigned by combining (such as through averaging) x sampling times. Let R be the total number of studied rooms and M the total number of APs. $(R+1)\times(M+1)$ is the radio map matrix where each row vector is along with its corresponding position except the first M

dimensional vector space, which contains the MAC addresses of the M visible access points denoted by $(AP)_{RM}$.

Step 2 (normalized rank transformation). This intermediate step is achieved by separately considering each row vector of the radio map. This consideration follows the described steps in Section 2 proceeding from (1) to (6).

Step 3 (model building and parameters adjustment). The model is built based on the new defined database and the transformed normalized rank values. The parameters of the learning method are tuned upon the obtained performance using the same device that is utilized to construct the radio map. The scrutinized criteria are the accuracy, precision, and the recall such that

$$Accuracy = \frac{TP + FN}{TP + FP + TN + FN}$$

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

$$Recall = \frac{TP}{TP + FN},$$

where TP, FP, TN, and FN are True Positive, False Positive, True Negative, and False Negative, respectively. Classification accuracy alone may hide details about the performance of classification model. It can also be misleading in the case of having more than two classes in the dataset. It gives the overall performance of the model considering all correct predictions divided by the total number of the datasets. However, it does not point out if all classes are predicted equally or whether some classes are neglected by the model. The formulations of the precision and the recall are used in our study to extract more information from the generated model. The precision gives the percentage of the correctly predicted instances among the total number of positive predictions, while the recall is the percentage of correctly predicted instances among the total number of positives.

```
(I) Dimensionality check
    (1)    for j = 1, 2, . . . , M do
    (2)      for i = 1, 2, . . . , N do
    (3)        if [(AP)_RM]^(j) = [(AP)_Ob]^(i)
    (4)          Keep (RSSI)^(i) of Ob vector at the position (j)
    (5)        elseif i = N & [(AP)_RM]^(j) ≠ [(AP)_Ob]^(i)
    (6)          Pad position (j) with zero value
    (7)    Loop
(II) Normalized rank transformation
      (i) Keep only non-zero positions and get the new Ob′ vector
      (ii) Rearrange RSSI value in ascending order.
      (iii) Assign transformed Normalized Rank value:
          (1)  Initialize the Normalized Rank (NR-Ob) vector
          (2)  for k = 2, 3, . . . , iterator do
          (3)    if (Ob′)_k = (Ob′)_{k−1}
          (4)      (NR-Ob)_k = Ψ_{k−1}
          (5)    else
          (6)      (NR-Ob)_k = Ψ_{k−1} + 1
          (7)  Loop
          (8)  λ = max{(NR-Ob)_{(1:iterator)}}
          (9)  Normalize NR-Ob vector by λ and remapping to M dimensional space
          (10) Consider the transformed normalized rank vector for positioning
(III) ML Classification
(IV) Return the final position
```

ALGORITHM 1: Proposed normalized rank transformation method during runtime phase.

3.2. The Online Stage. This part is fully detailed in Algorithm 1. Like the offline phase, this stage consists of three main steps: *Step 1:* dimensionality check; *Step 2:* normalized rank transformation; *Step 3:* ML classification and position estimation.

Step 1 (dimensionality check). During our database construction, considering M visible APs leads to the need of dimensionality check. This step checks each given observed vector at the location, which is denoted (Ob) and is to be identified. Let N be the total number of visible APs of (Ob) vector. Dimensionality transformation from N dimension space to M dimension space is established such that the RSSI value at i location in (Ob) vector corresponding to a specific AP at j location in $(AP)_{RM}$ vector should be stored at the same j location. However, for the unseen APs, their corresponding RSSI values are padded with zero to achieve the same dimension in both vectors. Furthermore, it allows pattern matching to take the same features into consideration during the location estimation process.

Step 2 (normalized rank transformation). The RSSI values are transformed to the normalized rank values by keeping the M dimensional space of the observed vector.

Step 3 (ML classification and position estimation). The transformed normalized rank vector (NR-Ob) is compared with the trained model to find the best match. The physical position of the model which has the best match in the new radio map will be labeled as the estimated position.

The same RSSI transformation has been applied in both offline and online stages. Several classifiers will be trained using the new radio map, generating the training model to validate our approach.

4. Classification Methods

In this section, we introduce the different methods adopted in our work. All the studied algorithms fall under the category of the supervised techniques.

4.1. Support Vector Machine. Support Vector Machine is one of the best off-the-shelf supervised learning algorithms, which is used for both classification and regression tasks. It has been originally developed for binary classification problems and further expanded to perform even multiclass classification tasks. It uses hyperplanes to define decision boundaries separating data points of different classes in the case of linearly separable data. In addition, SVMs can efficiently perform a nonlinear classification by using kernel trick. In this case, original data are mapped into a high-dimensional, or even infinite-dimensional, feature space [31, 32].

SVM aims to construct a hyperplane with the maximal margin between different classes. In most cases, data are not perfectly linearly separable, which makes the separating hyperplane susceptible to outliers. Therefore, a restricted number of misclassifications should be tolerated around the margins. The resulting optimization problem for SVMs,

where a violation of the constraints is penalized, depends on the regularization norm considered.

4.1.1. L1 Regularization Norm for SVM. The L1 norm defines a Support Vector Machine with a linear sum of the slack variable to make the hyperplanes less sensitive to outliers; it is written as

$$\min_{w,\xi,b} \quad J(w,\xi) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i \tag{8}$$

$$\text{Subject to:} \quad y_i\left(w^T\phi(x_i)+b\right) \geq 1-\xi_i \tag{9}$$

$$i = 1,\ldots,N, \ \xi_i \geq 0, \ i = 1,\ldots,N$$

$$y_i = \text{sign}\left(w^T\phi(x_i)+b\right), \tag{10}$$

where J is the cost function, w is the weight vector, and C is the positive regularization constant which defines the tradeoff between complexity and proportion of nonseparable samples. The problem formulation in (8) and (9) refers to the primal optimization problem. Introducing the Lagrange multipliers $\alpha_i \geq 0$, we obtain the following dual problem:

$$\max_{\alpha} \quad W(\alpha)$$

$$= \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{N} y_i y_j \alpha_i \alpha_j \phi(x_i)^T \phi(x_j) \tag{11}$$

$$\text{Subject to:} \quad \sum_{i=1}^{N} y_i \alpha_i = 0, \quad 0 \leq \alpha_i \leq C.$$

The optimal hyperplane is the one which maximizes the margin, and the optimal values of w and b are found by solving a constrained minimization problem using Lagrange multipliers. The function that solves the quadratic programming problem, and with the use of the positive definite mapping function that satisfies Mercer's condition, is such that

$$k(x,x_i) = \phi(x)^T \phi(x_i). \tag{12}$$

It also satisfies the Karush-Kuhn-Tucker (KKT) conditions.

The decision function can be expressed as

$$y_i = \text{sign}\left(\sum_{i=1}^{N}\alpha_i y_i k(x,x_i)+b\right). \tag{13}$$

The classification accuracy produced by SVMs may show variations depending on the choice of the kernel function and its parameters. Various types of kernels can be chosen:

(i) Linear:

$$K(x,x_i) = x^T x_i. \tag{14}$$

(ii) Polynomial of degree d:

$$K(x,x_i) = \left(\gamma + x^T x_i\right)^d, \quad \gamma \geq 0. \tag{15}$$

(iii) Radial basis function (RBF):

$$K(x,x_i) = \exp\left(-\frac{\|x-x_i\|_2^2}{\delta^2}\right). \tag{16}$$

(iv) Sigmoid:

$$K(x,x_i) = \tanh\left(k_1 \cdot x^T x_i + k_2\right)^d. \tag{17}$$

The KKT condition is given by

$$\alpha_i\left(y_i\left(w^T x_i + b\right) - 1 + \xi_i\right) = 0$$

$$b_i \xi_i = (C - \alpha_i)\xi_i = 0. \tag{18}$$

α_i are Lagrange multipliers and the value ξ_i indicates the distance of x_i with respect to the decision boundary since that

(i) if $\alpha_i = 0$, then $\xi_i = 0$; therefore x_i is correctly classified and lies outside the margin;

(ii) $0 < \alpha_i < C$; then $(y_i(w^T x_i + b) - 1 + \xi_i) = 0$ and $\xi_i = 0$; thus $y_i(w^T x_i + b) = 1$ and x_i is the support vector;

(iii) $\alpha_i = C$; then $(y_i(w^T x_i + b) - 1 + \xi_i) = 0$ and $\xi_i \geq 0$; therefore x_i is a bounded support vector; in the case $0 \leq \xi_i < 1$, x_i is correctly classified; however, for $\xi_i \geq 1$, x_i is misclassified.

4.1.2. L2 Regularization Norm for SVM. The L2 norm defines a Support Vector Machine which uses the square sum of the slack variable in the objective function. The considered optimization problem is as follows:

$$\min_{w,\xi,b} \quad J(w,\xi) = \frac{1}{2}\|w\|^2 + \frac{C}{2}\sum_{i=1}^{N}\xi_i^2$$

$$\text{Subject to:} \quad y_i\left(w^T x_i + b\right) \geq 1 - \xi_i \tag{19}$$

$$i = 1,\ldots,N, \ \xi_i \geq 0, \ i = 1,\ldots,N.$$

The dual problem is expressed by introducing the Lagrange multipliers α_i:

$$\max_{\alpha} \quad W(\alpha)$$

$$= \sum_{i=1}^{N}\alpha_i$$

$$- \frac{1}{2}\sum_{i,j=1}^{N} y_i y_j \alpha_i \alpha_j \left(k(x,x_i)+\frac{\delta_{ij}}{C}\right) \tag{20}$$

$$\text{Subject to:} \quad \sum_{i=1}^{N} y_i \alpha_i = 0, \quad \alpha_i \geq 0 \text{ for } i = 1,\ldots,N,$$

where δ_{ij} is Kronecker's delta function, in which $\delta_{ij} = 1$, for $i = j$, and 0 otherwise.

The KKT conditions in this case are given by

$$y_i\left(\sum_{j=1}^{N}\alpha_j y_j\left(k(x,x_i)+\frac{\delta_{ij}}{c}\right)+b\right) - 1 = 0. \tag{21}$$

4.2. KNN. K-Nearest Neighbor algorithm is a nonparametric supervised classifier in which the target label is predicted by finding the nearest neighbor class, considering the majority vote of its K neighbors. In the case of $K = 1$, the target is simply assigned to the class of its nearest neighbor. The closest class in this work is identified using the Euclidean distance:

$$D_{ij} = \sqrt{\sum_{k=1}^{M} \left(X_{ik} - X_{jk} \right)^2}, \tag{22}$$

where D_{ij} is the distance between the observed point and each signature vector in the radio map. X_i corresponds to the runtime fingerprint, X_j is the offline fingerprint, and M corresponds to the dimension of the RSSI Vector. The best choice of the calibration points K selected in this work has been set equal to one, since it generated better results.

4.3. Naïve Bayes. Naïve Bayes is a probabilistic classifier based on applying Bayes theorem with strong naïve independence assumptions between the features. Each data instance generates a tuple X of attribute values $\langle x_1, x_2, \ldots, x_M \rangle$. The identification of the corresponding label from a finite set of labels R consists on maximum a posteriori (MAP) estimation. The theorem states the following relationship:

$$P\left(r \mid x_1, x_2, \ldots, x_M \right) = \frac{p\left(r \right) P\left(x_1, x_2, \ldots, x_M \mid r \right)}{P\left(x_1, x_2, \ldots, x_M \right)}, \tag{23}$$

since $P(x_1, x_2, \ldots, x_M)$ is constant given the input set:

$$r = \arg\max p\left(r \right) P\left(x_1, x_2, \ldots, x_M \mid r \right)$$
$$r = \arg\max p\left(r \right) \prod_{i=1}^{M} P\left(x_i \mid r \right), \tag{24}$$

where r is the estimated room, which is predicted given the transformed rank values from M APs, $p(r)$ is the prior probability of the class "r," and $P(x_i \mid r)$ corresponds to the likelihood. Both terms are estimated from the training data. We adopt the implementation of Naïve Bayes considering the fact that continuous variable with each class is distributed according to a Gaussian distribution:

$$P\left(x_i = v \mid r \right) = \frac{1}{\sqrt{2\pi\delta_r^2}} e^{-(v-u_r)^2/2\delta_c^2}. \tag{25}$$

4.4. Random Forest. Random Forest (RF) is a classifier in which a multitude of decision trees are generated. The RF chooses the tree which has the highest votes after their classification results. The most occurring class number in the output of the decision trees is the final output of the RF classifier. A recursive process in which the input dataset is composed of smaller subsets allows the training of each decision tree. This process continues until all the tree nodes reach the similar output targets. The Random Forest classifier takes weights based on the input as a parameter that resembles the number of the decision trees [18].

4.5. Artificial Neural Network. Artificial Neural Network (ANN) is one of the most effective models in ML. It has been inspired by the biological neural networks in the human brain. It is made of several units or neurons of the following form:

$$H_j = \sigma \left(b + \sum_{i=1}^{N} w_{ij} x_i \right), \tag{26}$$

where σ is a nonlinear activation function, b is the bias term, and w_{ij} are the associated weights to the column vector x (corresponding either to the input data or the preceding layer).

In this paper, our selected activation function is the sigmoid function as in the following [20]:

$$\sigma\left(x \right) = \frac{1}{1 + e^{-x}}. \tag{27}$$

The nodes in ANN are structured into successive layers: input layer which corresponds to the input data, hidden layers, and output layer. The required number of hidden layers depends on the nonlinearity of the relation between input and output. Backpropagation algorithm is used to adjust weights and bias values of the edges and to minimize the loss function.

5. Experimental Results and Analysis

Our experiments were conducted in Metro City shopping mall in Shanghai and we considered about 88 different shops during the calibration phase. Figure 2 shows one floor plan of this shopping mall. Only one device has been used to build the radio map based on the collected RSSI measurements. The total visible M APs considered herein are equal to 185. Samsung Galaxy Note 2 mobile phone is considered as the reference device.

During the testing phase, sixteen different devices have been utilized to investigate hardware variance effect on localization accuracy. These devices recorded the signal strength from the available APs with their corresponding MAC addresses at the same locations in 8 shops. In each shop, sixty samples have been recorded, with a total of 480 samples based on each single device. Different database sizes denoted by S (corresponding to the number of assigned observations to each position) have been utilized to investigate how much we need to know ahead of time about what is being learned. This analysis aims to achieve an effective learning and correctly predict the position, for new observations unseen before. Moreover, the accuracy based on a single observation, as well as the needed fused data, is studied. The performance of the proposed method is evaluated through extensive experiments, and the obtained results based on SVM are compared with other well-known and widely used ML algorithms in indoor localization. In this section, we assess the performance of the built model based on a reference device to diverse devices. In the first part, we investigate database definition based on the collected RSSI value. It is implemented by assigning a varied set of observations to each single location and maintaining the absolute RSSI value during the runtime
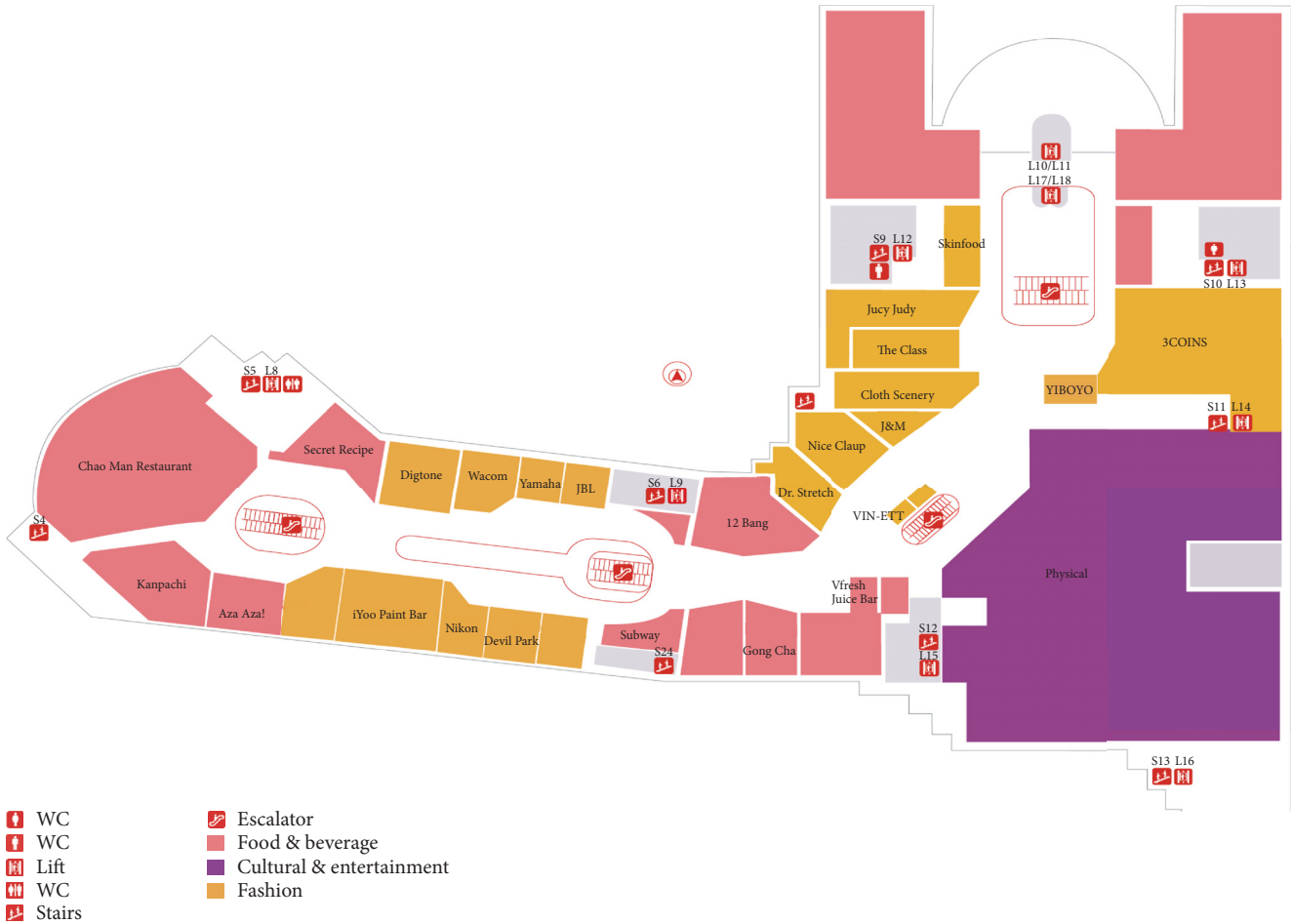
FIGURE 2: 3rd floor plan of Metro City shopping mall.

phase. The predicted location is calculated by matching the sample points on the radio map, with the RSSI fingerprint closest to the tracking device.

The linear kernel of SVM based on risk minimization principle is adopted to make the decision about the current location. Seeing that, for the most part of analysis, the number of input variables exceeds the number of examples, which makes the linear separation well suited based on Cover's theorem [33], same kernel parameters are considered during the whole analysis to fairly compare the obtained results. Moreover, we checked the efficiency of the rank transformation without any normalization. In addition, to evaluate the effectiveness of the proposed method, we considered about two ways of normalization. The first technique is the fully normalized rank, considering the total stored input in the database. However, the second one is the vector normalized rank method in which the normalization factor λ is taking the value as described in Section 2.

5.1. Experiments Based on RSSI Values. Figure 3 shows the positioning accuracy with different S and the influence of changing the tracking device when using RSSI values in both online and offline phase. The vertical axis shows the percentage of correctly predicted positions. We look first at the achieved accuracy in the case of using the same training device in the positioning phase. The good and steady performance during this test can be seen with high precision to estimate the location, where only a few samples are enough to attain the maximum accuracy. Almost all locations are perfectly estimated in this case.

However, in the case of tracking device different from the training device during the runtime phase, it works badly in half of the studied cases. Besides, this observation is much more remarkable if the number of fingerprint observations in the same location is limited or very small. Although the achieved correct rate for the remaining smartphones turns around 100% well-estimated location, only some of them can be distinguished in the figure. Unsteady curves are noticeable and increasing the number of the encoded fingerprints based on RSSI values at each location on the radio map does not always improve the performance of the applied method. This outcome proves the degradation pattern caused by the change of the utilized device to record the signal strength. This is due to the fact that RSSI stored in the database diverges from the RSSI captured using another piece of equipment. This low estimation is coming from the different sensitivity of
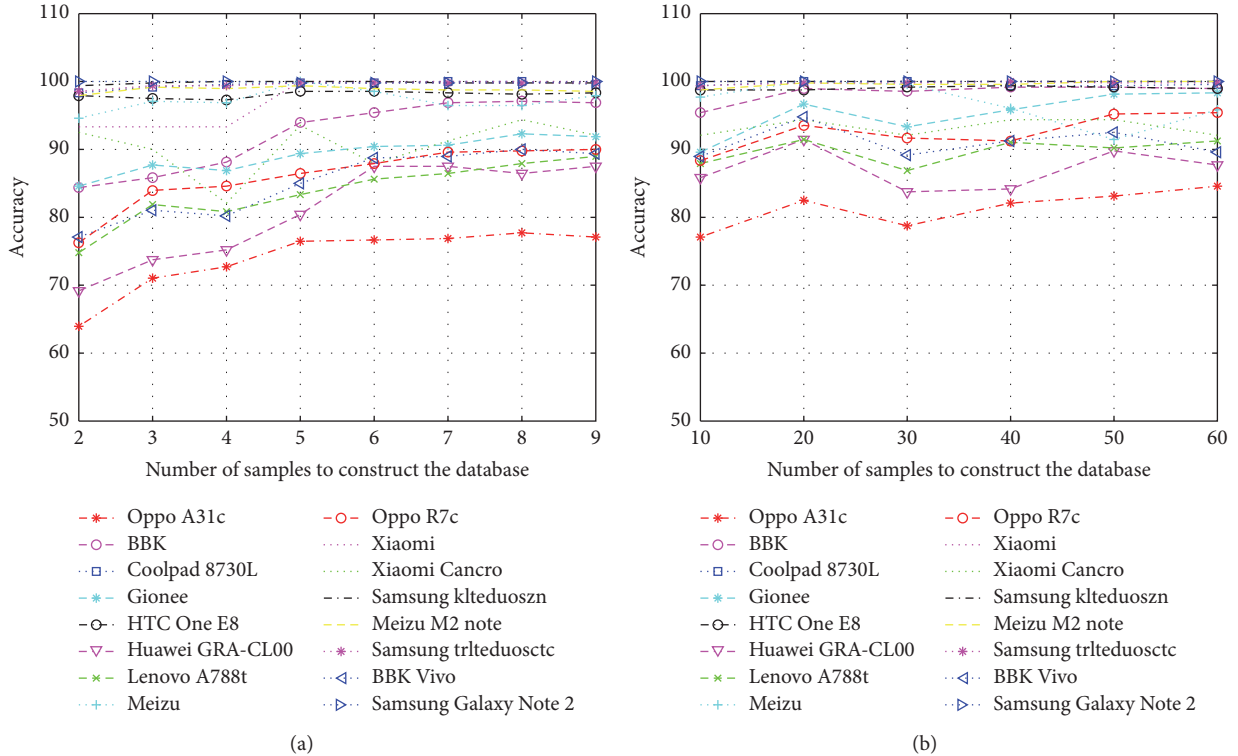
FIGURE 3: Localization accuracy using RSSI based on linear SVM. (a) shows the accuracy based on a few samples. (b) shows the accuracy based on large samples.

manifold devices to the signal throughput, which is related to both antennas and packaging materials.

5.2. Experiments Based on the Rank Transformed Values. In this part, we have redone the previous tests based on the rank transformed values instead of exploiting the direct RSSI value to predict the location. The RSSI vector $P_{1:M} = \{S_1, S_2, \ldots, S_M\}$ is reformulated such that $\Psi_{1:M} = \{\Psi_1, \Psi_2, \ldots, \Psi_M\}$, where Ψ_i satisfies (4) and the normalization factor is initially ignored. Figure 4 illustrates the location accuracy by changing the number of associated observations to each Reference Point (RP) using multiple devices. A prominent improvement is noticeable using the rank transformed value and more stability is perceptible comparing to the previous results. The maximum correct rate estimation is much higher based on a few associated samples to the reference locations. The reached perfect predictions based on this test have not been attained based on RSSI analysis even when we considered the integrality of captured observations. The two exception cases of *Xiaomi* and *Oppo A31c* devices will be discussed in the next part.

Ranking the signal throughput of an access point at a specific location plays an essential role in delimiting the range of the interval in which the input variable is defined. It is additionally practical to broaden the application of a built model. Moreover, the prioritization of the most reliable input variables when assigning rank values strengthens the generalization of the model and the performance of the used method.

5.3. Experiments Based on the Normalized Rank Transformation Method. In order to evaluate the performance of the normalized rank transformation, we define two means of normalization. Figure 5 represents the performance of the learning method using the fully normalized rank transformation. The normalization factor λ takes the inverse value of M where M is the total visible APs considered in the radio map, while Figure 6 corresponds to the achieved results when λ satisfies the condition of (5). It is noteworthy that both ways can provide more accurate results with some meaningful differences.

It appears that normalizing based on a fixed value needs more samples to attain a good prediction if compared to the second normalization side by side. It can be seen that, for the greatest part, at least 6 samples are required to reach an approximate 100% accurate prediction. The recorded RSSI values using *Xiaomi* device have been compared with those obtained with the remaining devices, where an unpredicted variability has been registered. We notice that some samples are very similar for the same environment. Meanwhile, it could exhibit a remarkable inconsistency in the recorded throughput at the same position. It seems that this mobile phone is very sensitive to changes and to the stability of the studied area which explains the registered uncertainty. The maximum percentage of 87.5% achieved by *Oppo A31c* is further interpreted based on the confusion matrix.

Nonetheless, the second followed normalization based on the highest assigned transformed rank value is rapidly converging to the maximum accuracy, except for the special
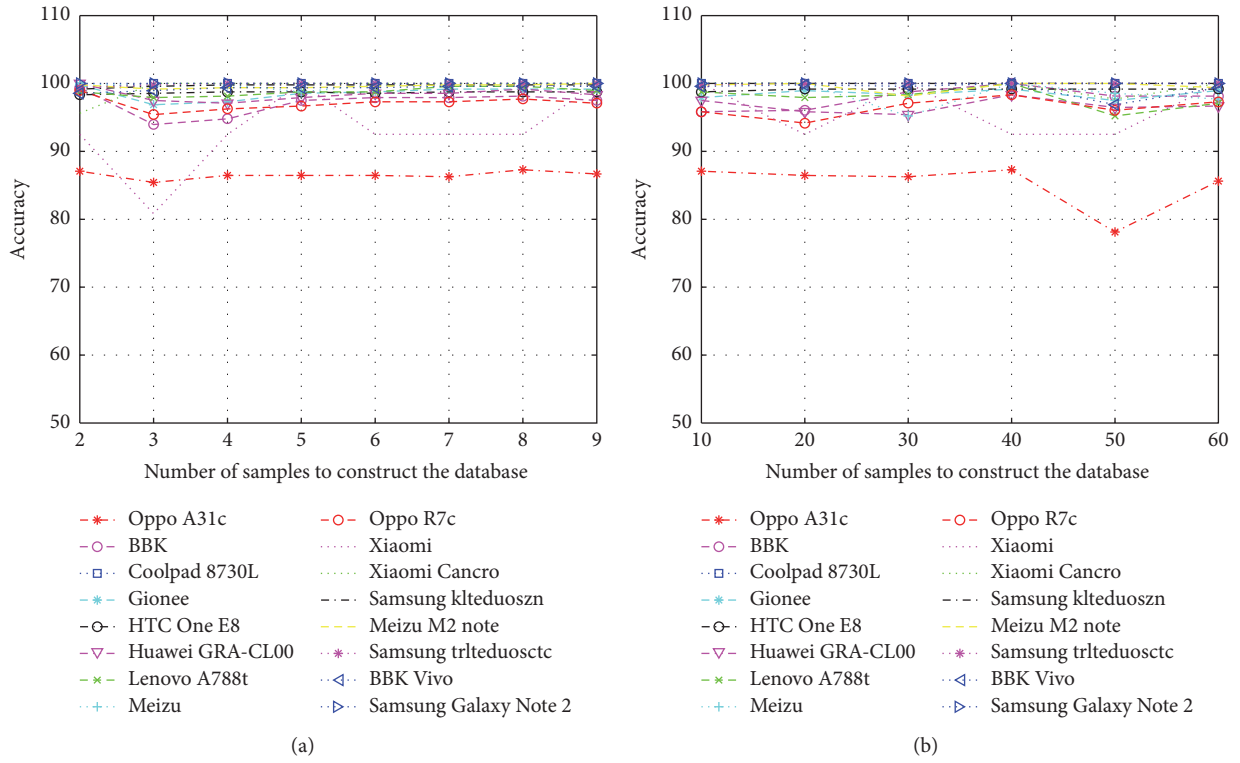
FIGURE 4: Localization accuracy using the rank based on linear SVM. (a) shows the accuracy based on a few samples. (b) shows the accuracy based on large samples.
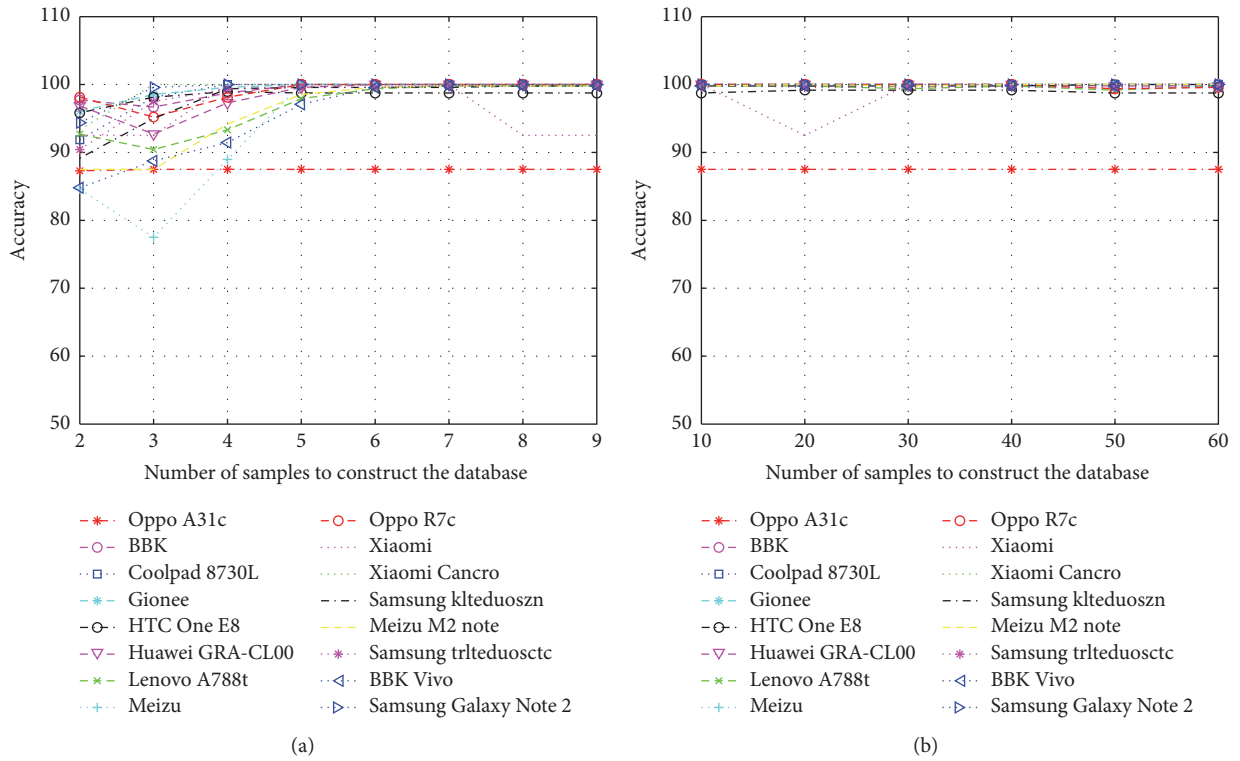

FIGURE 5: Localization accuracy using the fully normalized rank based on linear SVM. (a) shows the accuracy based on a few samples. (b) shows the accuracy based on large samples.
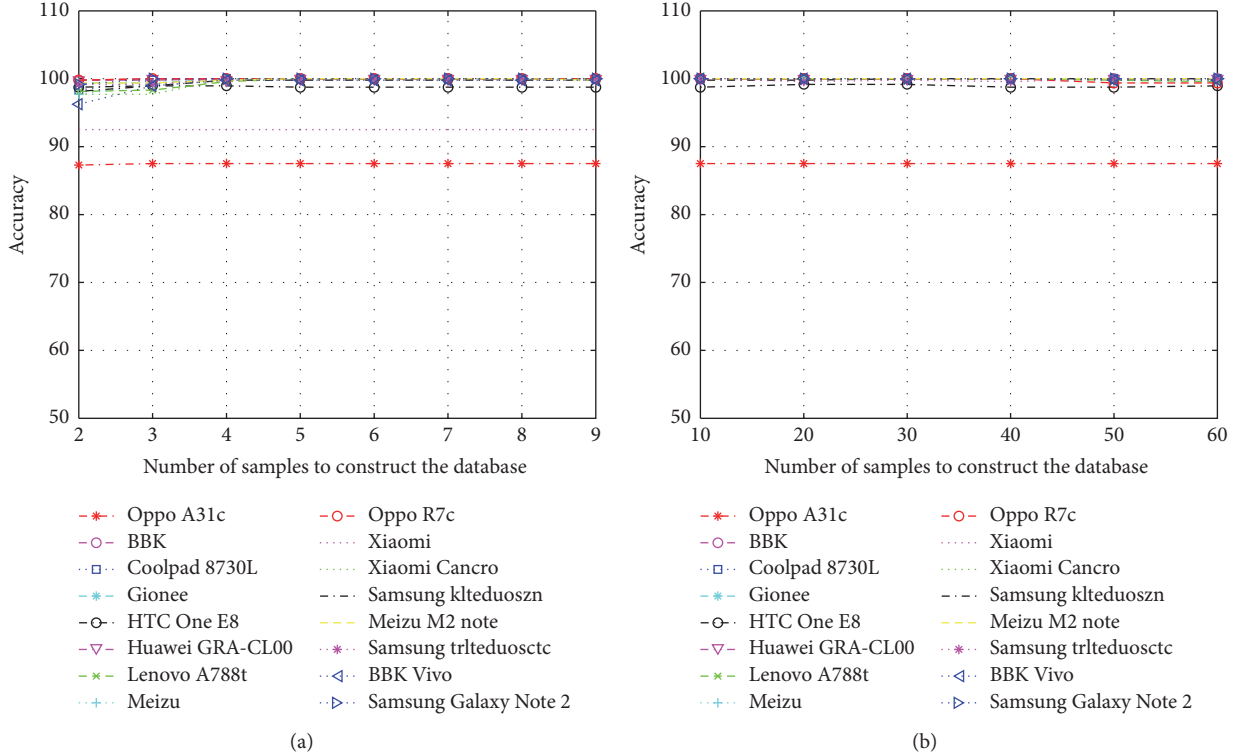
FIGURE 6: Localization accuracy using the vector normalized rank based on linear SVM. (a) shows the accuracy based on a few samples. (b) shows the accuracy based on large samples.

case of *Xiaomi* device that reached its maximum until the consideration of the 10th sample. This approach allows considering the varying range of the allocated tags in the new formulated vectors at one position, and consequently the difference between vectors dimensionalities. In contrast to the previous normalization, in the case of $\lambda = 1/\Psi_M$, it does not only perform as a scaling factor but also puts the transformed values of both real location fingerprint vector and the runtime vector as close as possible.

5.4. Experiments Based on a Unique Set of Observations.
The experiments in this scenario consider about one sample designating a special Reference Point. In the test, the stored fingerprints in the radio map are attained by averaging the whole recorded dataset. Figure 7 illustrates the percentage of the correct location determination based on a small size database using the linear SVM classification. Figures 7(a), 7(b), and 7(c) are, respectively, obtained by accounting for the absolute RSSI value, the inverse rank value, and the rank value. Finally, Figure 7(d) shows the performance of the learning method by applying the proposed method. The aim of comparing between Figures 7(b) and 7(c) is to demonstrate the importance of considering the throughput signal of an input variable in the right direction. As notable in Figure 7(b), a wrong direction leads to decreasing the accuracy comparing with the reference Figure 7(a) based on RSSI value.

On the one hand, the allocation of tags in the right order results in providing the APs with a high received

signal strength by high ranks, while the less reliable ones are assigned the low values. On the other hand, this will result in the prioritization of the principal features by according minor consideration to the APs with low signal strength when building the support vectors. This is due to the fact that the decision boundaries of the support vectors of SVM are always dominated by the large quantities. The built model, in this case, is much fitting the needs of this study than the inverse rank transformation. Moreover, a significant improvement is apparent by implementing the proposed solution with no less than 98.75% accurate estimation in 93.75% of the tested cases and 100% accuracy in 56.25% of cases.

Now turning attention to *Xiaomi* device, it turns out that fusing data reduces the effect of the large variation in the registered signal strength and enhances the prediction performance. We attempted to investigate the volume of needed data to be fused to consider one fingerprint for a given Reference Point (RP). From Figure 8, it appears that both ways of normalization herein work better than just applying the rank transformation with more stable prediction. Quite similar performances are seen in Figures 8(b) and 8(c) increasing by the rise of combined information, comparing with Figure 8(a).

5.5. Algorithms Comparison.
This section is dedicated to the comparison and the evaluation of the effectiveness of SVM among multiple ML techniques within a single experimental environment. To validate our proposed method, we

(1) Oppo A31c   (9) Oppo R7c
(2) BBK   (10) Xiaomi
(3) Coolpad 8730L   (11) Xiaomi Cancro
(4) Gionee   (12) Samsung klteduoszn
(5) HTC One E8   (13) Meizu M2 note
(6) Huawei GRA-CL00   (14) Samsung trlteduosctc
(7) Lenovo A788t   (15) BBK Vivo
(8) Meizu   (16) Samsung Galaxy Note 2

(a)

(1) Oppo A31c   (9) Oppo R7c
(2) BBK   (10) Xiaomi
(3) Coolpad 8730L   (11) Xiaomi Cancro
(4) Gionee   (12) Samsung klteduoszn
(5) HTC One E8   (13) Meizu M2 note
(6) Huawei GRA-CL00   (14) Samsung trlteduosctc
(7) Lenovo A788t   (15) BBK Vivo
(8) Meizu   (16) Samsung Galaxy Note 2

(b) #

(1) Oppo A31c   (9) Oppo R7c
(2) BBK   (10) Xiaomi
(3) Coolpad 8730L   (11) Xiaomi Cancro
(4) Gionee   (12) Samsung klteduoszn
(5) HTC One E8   (13) Meizu M2 note
(6) Huawei GRA-CL00   (14) Samsung trlteduosctc
(7) Lenovo A788t   (15) BBK Vivo
(8) Meizu   (16) Samsung Galaxy Note 2

(c)

(1) Oppo A31c   (9) Oppo R7c
(2) BBK   (10) Xiaomi
(3) Coolpad 8730L   (11) Xiaomi Cancro
(4) Gionee   (12) Samsung klteduoszn
(5) HTC One E8   (13) Meizu M2 note
(6) Huawei GRA-CL00   (14) Samsung trlteduosctc
(7) Lenovo A788t   (15) BBK Vivo
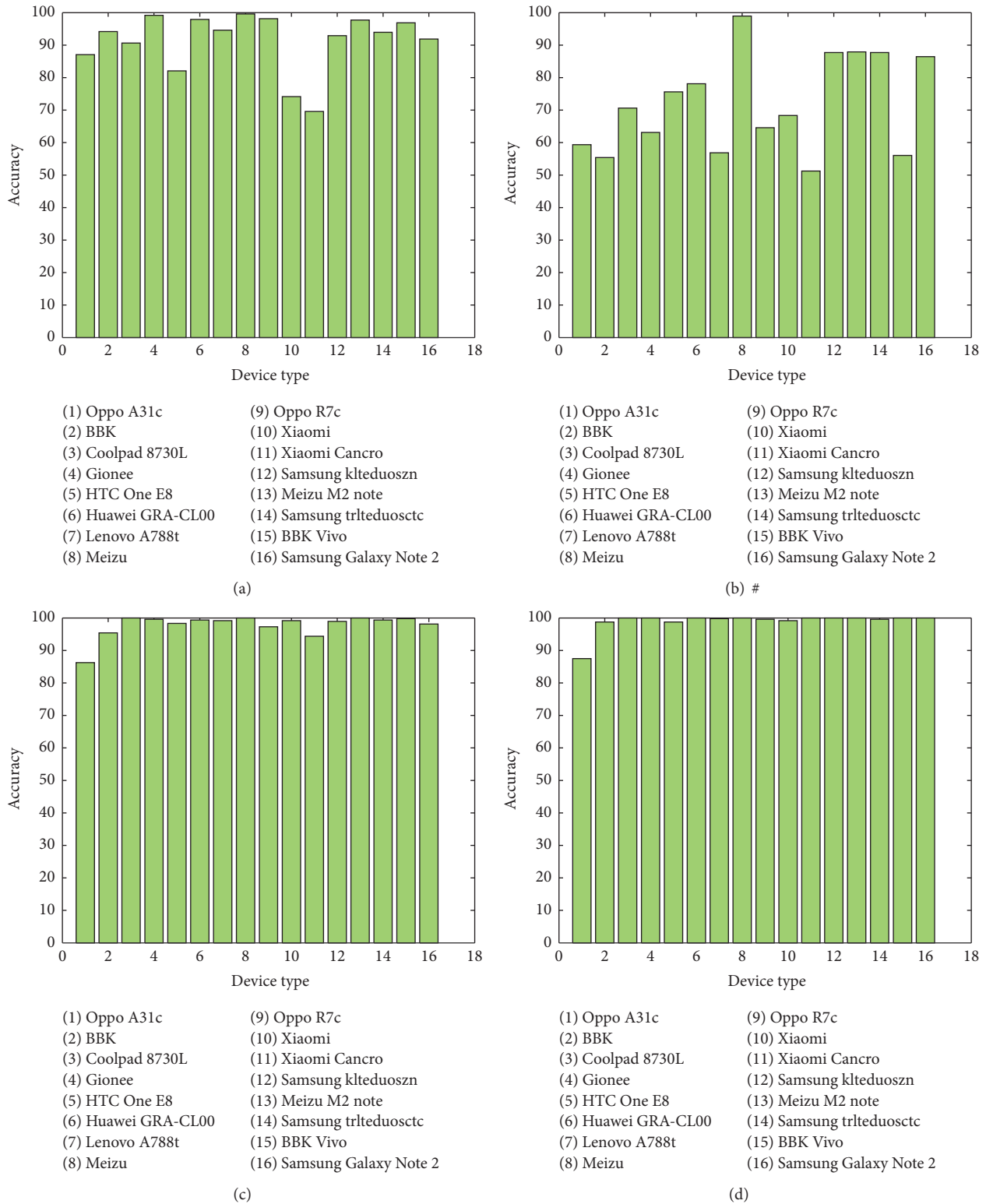(8) Meizu   (16) Samsung Galaxy Note 2

(d)

FIGURE 7: Localization accuracy comparison in the case of unique exemplification per location based on linear SVM. (a) shows the attained accuracy based on RSSI value. (b) shows the accuracy in the case of applying the inverse rank value. (c) shows the accuracy in the case of applying the rank transformation value. (d) Accuracy based on the normalized rank transformation. (#) The inverse rank is the converted absolute RSSI vector by considering the opposite direction of (3).
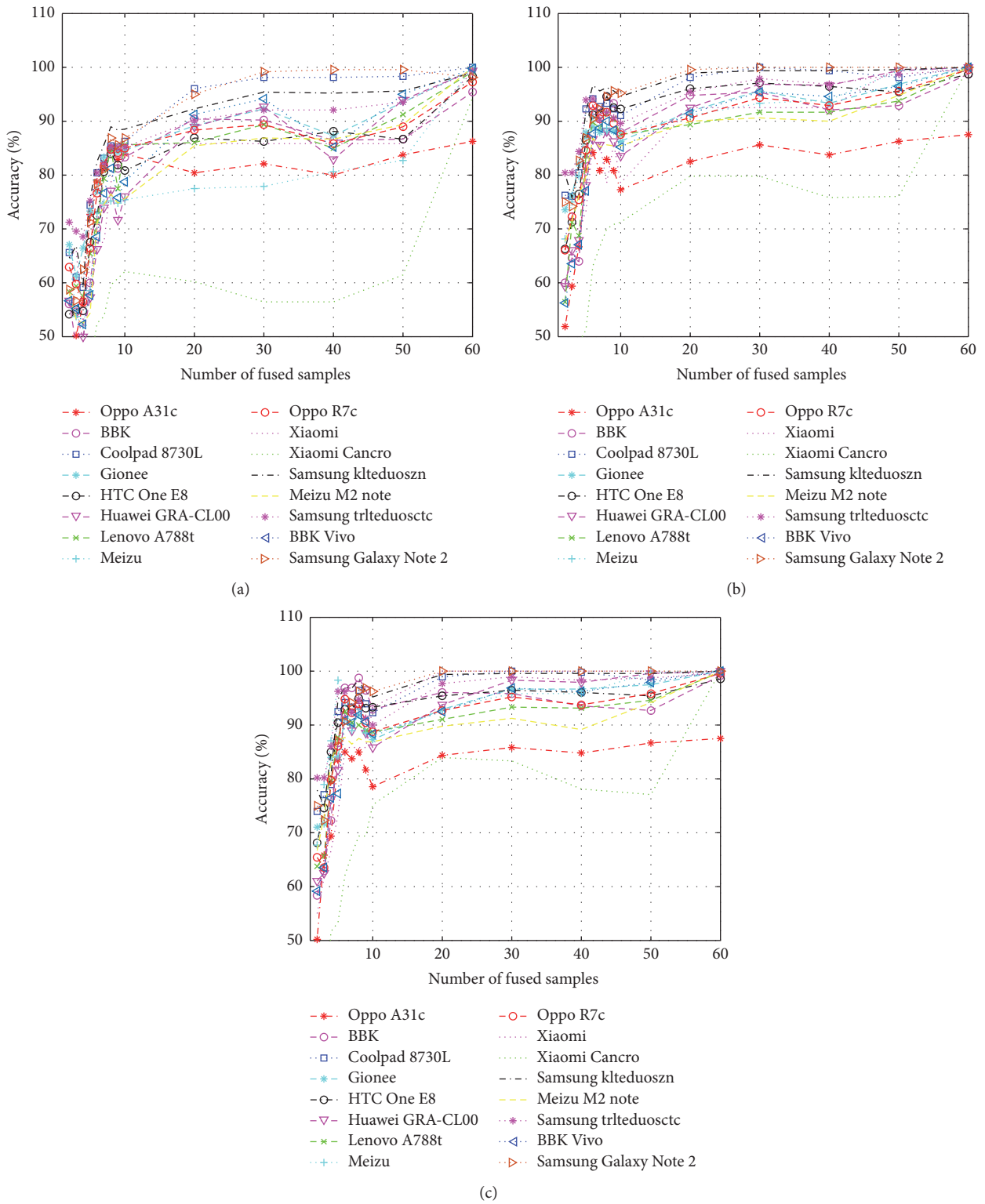
(a)

(b)

(c)

FIGURE 8: Localization accuracy in the case of unique exemplification per location based on linear SVM considering different data fusion size. (a) shows the attained accuracy based on the rank value. (b) shows the accuracy in the case of applying the proposed rank transformation. (c) shows the accuracy in the case of applying the fully normalized rank transformation.

TABLE 1: Percentage of accurately estimated locations based on linear SVM adopting multiple combinations of constraints violation along with comparison between RBF kernel and regularized Logistic Regression; the radio map is defined upon RSSI values.

Devices	Classification type						
	L2-R* L2-N (primal)	L2-R L2-N (dual)	L2-R L1-N (dual)	L2 LR (primal)	L1-R L2-N**	L1 LR***	RBF kernel
Oppo A31c	**87.08%**	**85.42%**	**85.42%**	**86.04%**	**52.29%**	**58.75%**	**37.29%**
BBK	94.17%	95.20%	95.20%	95.42%	**61.45%**	**70.20%**	**35.00%**
Coolpad 8730L	90.62%	**89.78%**	**89.78%**	90.42%	**70.42%**	**76.04%**	**41.45%**
Gionee	99.17%	99.37%	99.37%	99.37%	**58.12%**	**73.12%**	**37.29%**
HTC One E8	**82.08%**	**81.45%**	**81.45%**	**81.87%**	**54.58%**	**54.17%**	**36.04%**
Huawei GRA-CL00	97.92%	98.12%	98.12%	98.33%	**63.75%**	**64.58%**	**28.75%**
Lenovo A788t	94.58%	94.78%	94.78%	95.00%	**51.04%**	**66.25%**	**26.67%**
Meizu	99.58%	99.58%	99.58%	99.58%	**88.54%**	**87.29%**	**52.70%**
Oppo R7c	98.12%	98.12%	98.12%	97.92%	**55.62%**	**73.33%**	**36.45%**
Xiaomi	**74.17%**	**74.17%**	**74.17%**	**74.17%**	**40.83%**	**63.95%**	**37.50%**
Xiaomi Cancro	**69.58%**	**68.54%**	**68.54%**	**69.58%**	**37.50%**	**42.29%**	**32.91%**
Samsung klteduoszn	92.92%	92.50%	92.50%	93.12%	**68.54%**	**63.33%**	**47.50%**
Meizu M2 note	97.71%	97.08%	97.08%	98.33%	**83.75%**	**79.37%**	**58.12%**
Samsung trlteduosctc	93.95%	93.75%	93.75%	93.96%	**72.50%**	**66.45%**	**41.67%**
BBK Vivo	96.87%	97.08%	97.08%	97.29%	**57.91%**	**67.92%**	**30.20%**
Standard device accuracy	91.87%	91.87%	91.87%	92.29%	**79.79%**	**78.95%**	**54.79%**
Recall (ratio)	0.92	0.92	0.92	0.97	0.79	0.79	0.55
Precision (ratio)	0.95	0.95	0.95	0.97	0.86	0.83	0.77

(∗) R stands for regularization; (∗∗) N stands for norm; (∗∗∗) LR: stands for logistic regression. We put in bold the percentages lower than 90% for easiness of observation.

TABLE 2: Algorithms comparison based on the percentage of the well-recognized positions; the radio map defined upon RSSI values.

Device type	Classification type				
	SVM L2-R L2-N	Artificial Neural Network	Naïve Bayes	Random Forest	KNN ($K = 1$)
Oppo A31c	**87.08%**	97.50%	**77.50%**	**59.58%**	**46.04%**
BBK	94.17%	**56.25%**	**33.75%**	**59.38%**	**46.67%**
Coolpad 8730L	90.62%	97.71%	**64.38%**	**80.00%**	**81.46%**
Gionee	99.17%	**74.17%**	**37.71%**	**69.38%**	**58.96%**
HTC One E8	**82.08%**	91.88%	**56.88%**	**62.71%**	**81.04%**
Huawei GRA-CL00	97.92%	93.33%	**46.67%**	**66.04%**	**73.12%**
Lenovo A788t	94.58%	**78.75%**	**35.83%**	**61.88%**	**53.33%**
Meizu	99.58%	99.79%	95.00%	**84.00%**	99.58%
Oppo R7c	98.12%	**62.92%**	**36.04%**	**63.75%**	**50.83%**
Xiaomi	**74.17%**	81.04%	**55.00%**	**56.46%**	78.12%
Xiaomi Cancro	**69.58%**	**63.75%**	**38.75%**	**44.17%**	**52.92%**
Samsung klteduoszn	92.92%	98.75%	**67.29%**	**74.00%**	87.08%
Meizu M2 note	97.71%	97.71%	**86.25%**	**74.38%**	94.37%
Samsung trlteduosctc	93.95%	97.29%	**65.83%**	**68.96%**	89.17%
BBK Vivo	96.87%	**75.00%**	**39.17%**	**75.00%**	**52.29%**
Standard device accuracy	91.87%	97.50%	**77.50%**	**84.79%**	87.08%
Recall (ratio)	0.92	0.97	0.77	0.85	0.87
Precision (ratio)	0.95	0.97	0.85	0.93	0.93

We put in bold the percentages lower than 90% for easiness of observation.

compared the resulting estimation of the considered ML techniques with their ground-truth locations. Taking into account the importance of using a common database and the same way of data definition as considered in the offline phase, the conducted analysis is partitioned to two main tests. The first part is devoted to the tests based on RSSI value, while the second one is devoted to the proposed normalized rank transformation.

5.5.1. Algorithms Comparison upon RSSI Value. Both Tables 1 and 2 have been obtained by considering the absolute RSSI value in the calibration and the positioning phase. Table 1 shows the influence on accuracy by changing the considered regularization norm, the selected kernel for SVM classifier, and using the regularized logistic regression. Herein, for SVM classifier, the linear kernel yielded better results than the RBF kernel which achieved only 54.79% correct prediction, by

TABLE 3: Percentage of accurately estimated locations based on linear SVM adopting multiple combinations of constraints violation along with comparison between RBF kernel and regularized Logistic Regression; the radio map is defined upon the proposed normalized rank transformation.

Device type	Classification type						
	L2-R* L2-N (primal)	L2-R L2-N (dual)	L2-R L1-N (dual)	L2 LR (primal)	L1-R L2-N**	L1 LR***	RBF kernel
Oppo A31c	**87.50%**	**87.50%**	**87.50%**	**87.50%**	**58.95%**	85.83%	**85.20%**
BBK	98.75%	98.75%	98.75%	99.17%	**63.12%**	95.00%	**81.46%**
Coolpad 8730L	100%	100%	100%	100%	96.87%	99.58%	99.79%
Gionee	100%	100%	100%	100%	**71.87%**	96.04%	94.79%
HTC One E8	98.75%	98.75%	98.75%	98.75%	**87.50%**	91.04%	95.00%
Huawei GRA-CL00	100%	100%	100%	100%	**74.37%**	98.12%	97.08%
Lenovo A788t	99.79%	99.79%	99.79%	100%	**74.37%**	96.87%	98.54%
Meizu	100%	100%	100%	100%	94.58%	100%	100%
Oppo R7c	99.58%	99.58%	99.58%	100%	**70.62%**	95.62%	**87.92%**
Xiaomi	99.17%	99.17%	99.17%	99.17%	**76.46%**	**80.20%**	**85.41%**
Xiaomi Cancro	100%	100%	100%	100%	**74.37%**	91.25%	**83.33%**
Samsung klteduoszn	100%	100%	100%	100%	**88.75%**	97.29%	98.75%
Meizu M2 note	100%	100%	100%	100%	91.25%	94.37%	100%
Samsung trlteduosctc	99.58%	99.58%	99.58%	99.58%	92.08%	95.83%	98.95%
BBK Vivo	100%	100%	100%	100%	**70.83%**	93.95%	98.33%
Standard device accuracy	100%	100%	100%	100%	94.79%	99.37%	97.50%
Recall (ratio)	1	1	1	1	0.94	0.99	0.97
Precision (ratio)	1	1	1	1	0.91	0.99	0.98

(∗) R stands for regularization; (∗∗) N stands for norm; (∗∗∗) LR stands for logistic regression. We put in bold the percentages lower than 90% for easiness of observation.

fitting the parameters from our testbed, which was expected as stated in Cover's theorem.

It is noticeable that recognition rates and the generalization ability of the test data by L2-SVM tend to be better than those by L1-SVM. Moreover, the differences in the performance of L2-SVM applying either the dual or primal optimization problem are imperceptible. This means that L2 regularization is estimating the violating variables more conservatively than L1 regularization avoiding overfitting issue of the training samples. L1 regularization SVM tends to pick only a few variables in the case of the existence of several highly correlated variables. In addition, the number of selected variables is upper bounded by the size of the training data. However, L2 norm keeps the correlated variable by shrinking their corresponding coefficients.

It is noteworthy that the L2 regularized logistic regression behaves quite similarly to L2 regularized support vector since both of them tend to maximize the margin as reducing the loss although their considered loss is different. It is perceptible in a few cases that L2 regularized logistic regression produces slightly more accurate prediction compared to L2 regularized SVM. This is due to the fact that logistic regression is modeling probabilities and therefore some errors are calculated even for correctly classified training examples while L2-SVM does not. SVM classifier is purely discriminative and designed to give a binary classification, while the regularized logistic regression estimates the a posteriori probability of class membership. In this work, we desire a binary classification and consequently identify the decision boundary directly

rather than estimate the probability of class membership. On this basis we will look at the results of L2 regularized support vector instead of the L2 regularized logistic regression on the upcoming comparison.

Analysis using RSSI value based on linear SVM gives an acceptable result in most cases. Nonetheless, the model generalization to manifold devices could facilely fail to achieve a very high precision; for instance, a low correct rate estimation has been registered with *Oppo A31c, HTC, Xiaomi,* and *Xiaomi Cancro* with, respectively, a rate of 87.08%, 82.08%, 74.17%, and 69.58%. Table 2 illustrates the performance of the optimized algorithms, in which it can be seen that the built model of Random Forest, KNN, and Naïve Bayes failed dramatically, notably with heterogeneous devices. Good performances using Neural Network and SVM are noteworthy; however, SVM outperforms the whole implemented ML algorithms.

5.5.2. Algorithms Comparison upon the Proposed Normalized Rank Transformation. Tables 3 and 4 show the results of the proposed normalized rank transformation in both calibration and positioning phase. We investigated its usefulness considering regularization across LR and SVM classifier for linear and RBF kernels as depicted in Table 3, as well as its suitability of operation based on several ML algorithms as presented in Table 4. In general, we noticed the same previous observation based on RSSI analysis regarding the regularization and kernel type. However, a significant improvement is perceived moving from the RSSI value consideration to the normalized

TABLE 4: Algorithms comparison based on the percentage of the well-recognized positions; the radio map is defined upon the proposed normalized rank transformation.

Device type	Classification type				
	SVM L2-R	Artificial Neural Network	Naïve Bayes	Random Forest	KNN ($K = 1$)
Oppo A31c	**87.50%**	**86.67%**	**59.38%**	**57.29%**	**54.79%**
BBK	98.75%	**89.58%**	**64.38%**	**47.08%**	**64.58%**
Coolpad 8730L	100%	100%	95.83%	**82.08%**	**86.67%**
Gionee	100%	93.54%	**62.08%**	**52.50%**	**67.29%**
HTC One E8	98.75%	98.54%	96.25%	**80.42%**	**83.12%**
Huawei GRA-CL00	100%	98.54%	**71.25%**	**70%**	**81.04%**
Lenovo A788t	99.79%	96.67%	**64.38%**	**60.83%**	**77.29%**
Meizu	100%	100%	100%	93.54%	99.37%
Oppo R7c	99.58%	93.54%	**66.88%**	**61.67%**	**65.41%**
Xiaomi	99.17%	**83.96%**	**86.04%**	**76.88%**	**80.83%**
Xiaomi Cancro	100%	93.54%	**78.75%**	**63.54%**	**48.12%**
Samsung klteduoszn	100%	99.79%	99.58%	**86.46%**	**87.50%**
Meizu M2 note	100%	98.75%	99.58%	**86.46%**	92.50%
Samsung trlteduosctc	99.58%	99.17%	99.79%	**85.42%**	**83.75%**
BBK Vivo	100%	96.46%	69.38%	63.96%	70.83%
Standard device accuracy	100%	100%	100%	**89.38%**	90.83%
Recall (ratio)	1	1	1	0.89	0.91
Precision (ratio)	1	1	1	0.94	0.94

We put in bold the percentages lower than 90% for easiness of observation.

TABLE 5: Confusion matrix of *OPPO A31c* device.

	Shop 1	Shop 2	Shop 3	Shop 4	Shop 5	Shop 6	Shop 7	Shop 8
Shop 1	60	0	0	0	0	0	0	0
Shop 2	0	60	0	0	0	0	0	0
Shop 3	0	0	60	0	0	0	0	0
Shop 4	0	0	0	60	0	0	0	0
Shop 5	0	0	0	0	60	0	0	0
Shop 6	0	0	0	0	0	60	0	0
Shop 7	0	0	0	0	0	0	60	0
Shop 8	60	0	0	0	0	0	0	0

rank transformation. The converted absolute RSSI values seem to be more stable and reliable rather than their initial values. Analyzing the performance of the aforementioned ML algorithms, the accuracy based on the same device is ameliorated while the generalization of the model to diversiform devices has been outperformed based on NR-SVM. This solution shows its significance improving the optimized accuracy of SVM around the decision boundaries, where 100% perfect prediction is achieved on the same device based linear L2-SVM. Moreover, no less than 98.75% well-estimated locations are attained for the remaining devices, except the special case of *Oppo A31c*.

It is interesting to further examine the positioning error percentile for the special case of *Oppo A31c* where only 87.5% correct estimation has been recorded. We have drawn its corresponding confusion matrix given in Table 5 to summarize the performance of the classification algorithm for this device. The vertical shop specification corresponds to the actual location while the horizontal ones are the predicted shops. The diagonal part of the matrix delineates the correctly classified instance where sixty samples are defined in each shop for evaluation. It is observed that both locations "shop 1" and "shop 8" are highly confused where all vectors of shop 8 were classified as location "shop 1." In the case of having a high similarity between RSSI vectors in different locations, taking into account neither the initial RSSI values nor their converted NR values can significantly distinguish between them and provide accurate estimation.

5.6. RSSI Transformations Comparison Based on SVM. Several studies have proposed methods to improve the robustness of positioning systems against device diversity. Calibration-free approaches have introduced several ways of RSSI reformulation. This performance comparison is aiming to show the effectiveness of the proposed NR-SVM compared with the Hyperbolic Location Fingerprinting (HLF), the Signal Strength Differences (SSD), and the DIFF method based on SVM. The difference of signal strength between pairs of APs, namely, DIFF, was proposed in [25] to reduce the effect of diversity in devices. The main difference between
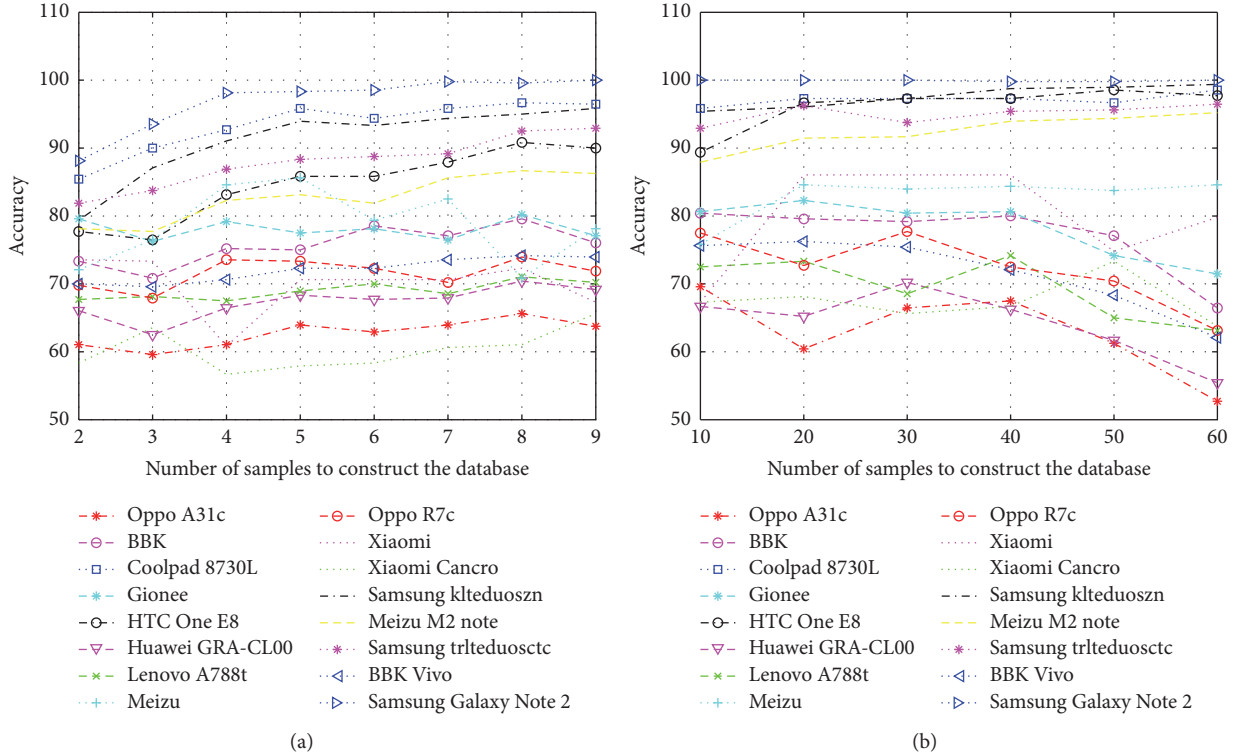
FIGURE 9: Achieved accuracy based on SSD-SVM. (a) shows the accuracy based on a few samples. (b) shows the accuracy based on large samples.

the fundamental fingerprinting method and DIFF is that the latest uses the difference between pairs of APs instead of the use of the absolute RSSI value. SSD [26] takes also the advantages of Signal Strength Differences; however, it selects only independent sets of DIFF. Moreover, the HLF [24] method uses the signal strength ratios between pairs of RSSIs as fingerprints.

We implemented the aforestated RSSI value transformations and estimated the location using SVM classifier. The results show that both DIFF (Figure 10) and HLF (Figure 11) conversions perform better than the initial RSSI values (Figure 3) either when using a few samples or by considering extrasite survey measurements to identify each location. It appears in our case of study that the SSD transformation achieves lower estimation in terms of accuracy comparing the estimated positions with their ground-truth locations. Focusing on the obtained curve using the same device versus a few samples Figure 9(a), the performance is clearly below that of the absolute RSSI values. The side effect of this approach is that the performance is not always better than the initial RSSI value or even worse with homogeneous devices notwithstanding the enlargement of training samples. To further investigate this observation, we carried out the test of HLF method based on independent sets only as shown in Figure 12. It turns out that this way of transformation based on independent sets yields a significant loss of some discriminative information which decreases the accuracy instead of improving it.

NR-SVM (Figure 6) performs the best and the experiments show a consistent result as well as a high precision comparing with three of the popular calibration-free techniques DIFF, SSD, and HLF based on SVM. The reached steady performance during the whole tests proves the feasibility of considering few samples even in the case of manifold devices. Consequently, it reduces the workload of the offline data training phase. Furthermore, it avoids the time-consuming site survey in which a large number of observations are collected to boost the reliability of the system.

6. Conclusion

In this paper, we have presented the normalized rank transformation method for a room level determination. The performance evaluation shows that is more efficient to redefine and reformulate the signal strength observed from the anchors adequately rather than to use the absolute RSSI values from an anchor node. The importance of ranking direction and the way of normalization have been compared. The normalization factor based on the highest assigned rank results in meaningful consideration of vector dimension yielding higher accuracy, and it is characterized by its fast convergence to the maximum possible accuracy.

We explore the efficiency of NR-SVM by formulating three of the most popular calibration-free transformations proposed in the literature. DIFF, SSD, and HLF are implemented based on SVM classifier. It has been concluded that
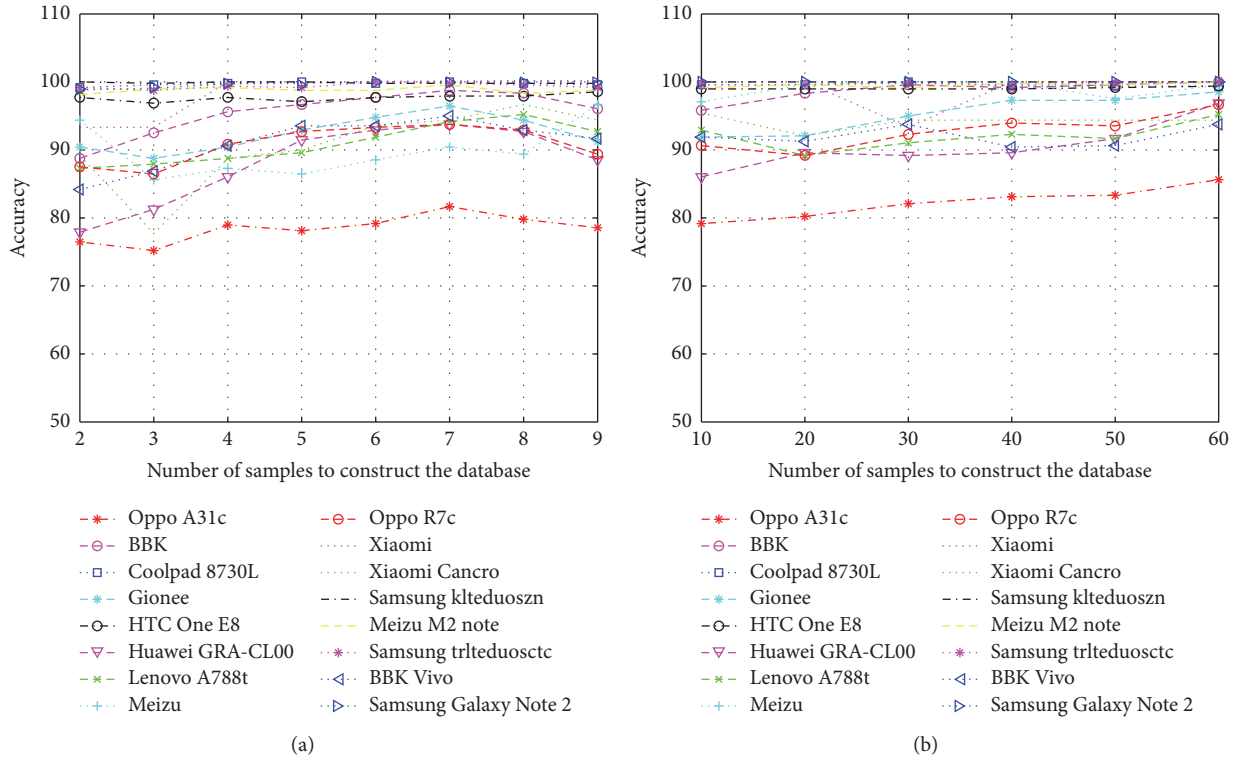
FIGURE 10: Achieved accuracy based on DIFF-SVM. (a) shows the accuracy based on a few samples. (b) shows the accuracy based on large samples.
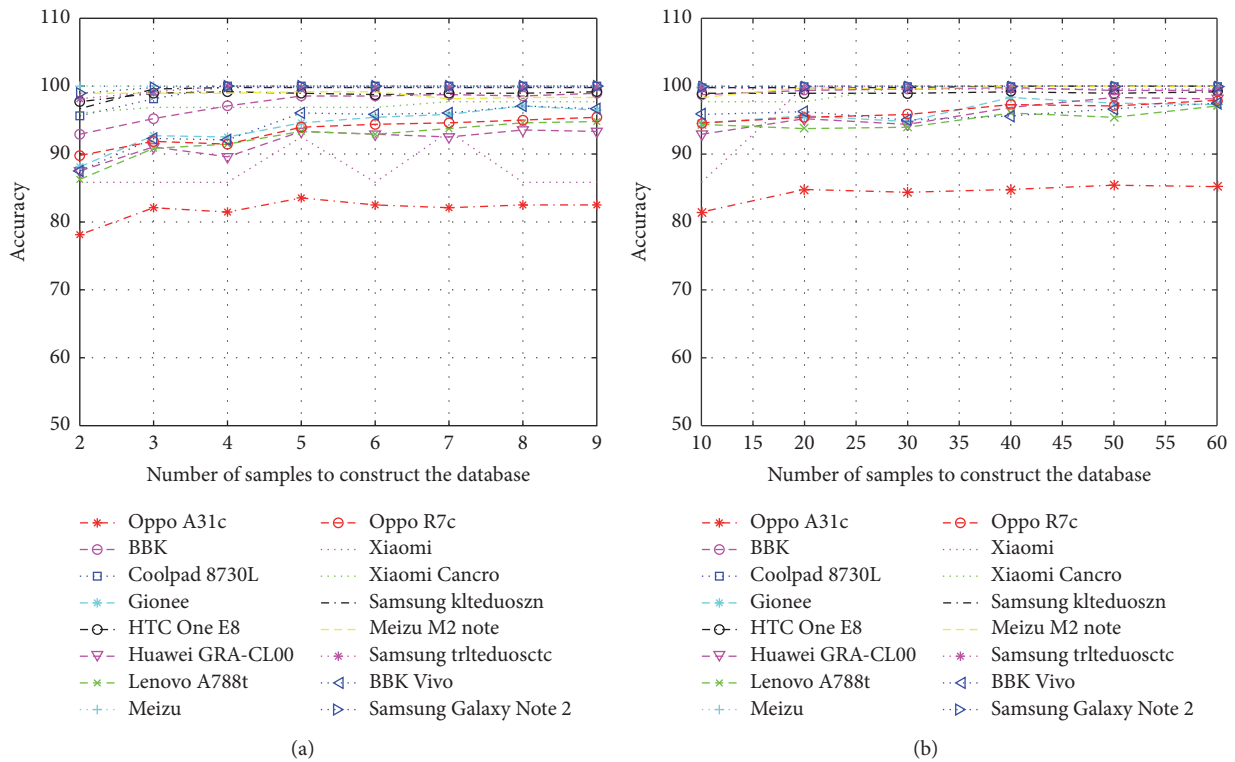

FIGURE 11: Achieved accuracy based on HLF-SVM. (a) shows the accuracy based on a few samples. (b) shows the accuracy based on large samples.
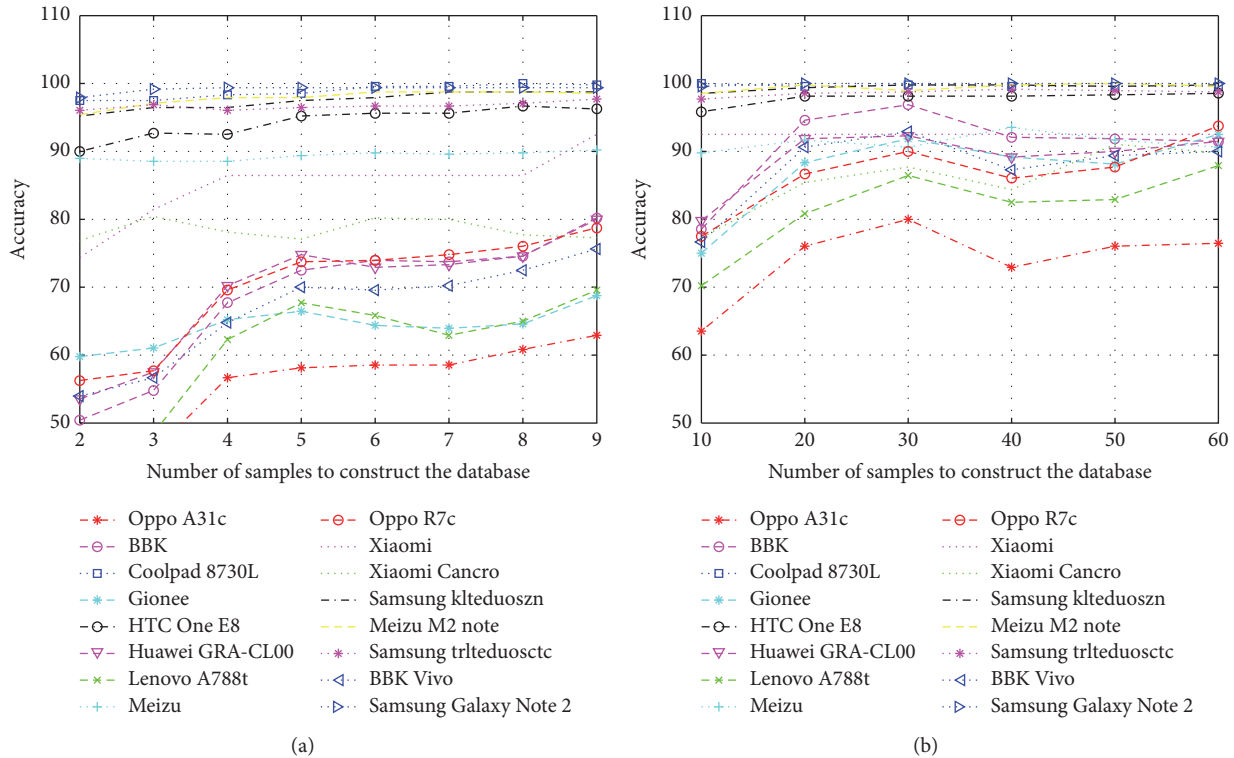
FIGURE 12: Achieved accuracy based on independent HLF-SVM. (a) shows the accuracy based on a few samples. (b) shows the accuracy based on large samples.

SSD transformation may result in a significant loss of some discriminative information and decreases the precision of the algorithm. However, both DIFF and HLF improved the performance compared to the fundamental fingerprinting method based on RSSI values. Moreover, NR-SVM has shown its distinguishability and reliability among the studied transformations.

Another contribution presented in this paper is the expansion of the application of the prebuild radio map upon a recorded information via a single device to manifold devices, taking into account the quality and stability over time of the resulting outcomes. Finally, we investigate and compare our method based on different machine learning algorithms and we showed that the linear L2 regularization norm of the Support Vector Machine outperformed Naïve Bayes, Random Forest, Artificial Neural Network, and KNN algorithms with moderate dataset in terms of accuracy, especially on handling device heterogeneity.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.
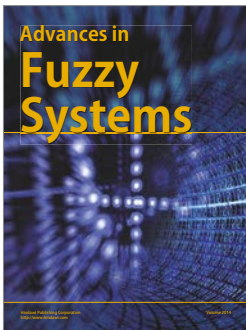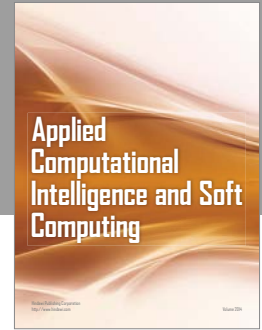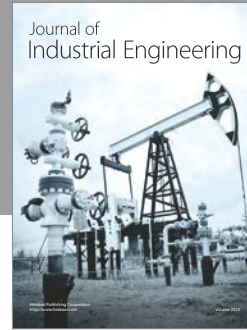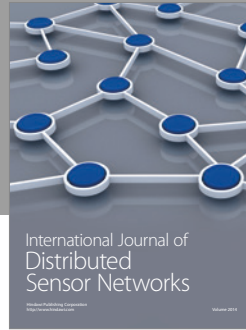
Acknowledgments

This work is supported by the Shanghai Science and Technology Committee under Grant 15511105100 and the

References

[1] J. Hightower, R. Want, and G. Borriello, "Spoton: an indoor 3D location sensing technology based on Rf signal strength," UW-CSE 00-02-02, Department of Computer Science and Engineering, University of Washington, Seattle, Wash, USA, 2000.

[2] P. Bahl and V. N. Padmanabhan, "Radar: an in-building RF-based user location and tracking system," in *Proceedings of the 19th Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE INFOCOM '00)*, pp. 775–784, Tel Aviv, Israel, March 2000.

[3] R. Chen, L. Pei, J. Liu, and H. Leppäkoski, "WLAN and blue-tooth positioning in smart phones," in *Ubiquitous Positioning and Mobile Location-Based Services in Smart Phones*, pp. 44–68, 2012.

[4] L. Pei, R. Chen, J. Liu, T. Tenhunen, H. Kuusniemi, and Y. Chen, "Inquiry-based bluetooth indoor positioning via RSSI probability distributions," in *Proceedings of the 2nd International Conference on Advances in Satellite and Space Communications (SPACOMM '10)*, pp. 151–156, Athens, Greece, June 2010.

[5] L. Pei, R. Chen, J. Liu, H. Kuusniemi, T. Tenhunen, and Y. Chen, "Using inquiry-based Bluetooth RSSI probability distributions for indoor positioning," *Global Positioning Systems*, vol. 9, no. 2, pp. 122–130, 2010.

[6] G. Gomes and S. Helena, "Indoor location system using ZigBee technology," in *Proceedings of the 3rd International Conference on Sensor Technologies and Applications (SENSORCOMM '09)*, pp. 152–157, Athens, Greece, June 2009.

[7] K. Pahlavan, F. O. Akgül, M. Heidari, A. Hatami, J. M. Elwell, and R. D. Tingley, "Indoor geolocation in the absence of direct path," *IEEE Wireless Communications*, vol. 13, no. 6, pp. 50–58, 2006.

[8] J. MacHaj, P. Brida, and J. Benikovsky, "Using GSM signals for fingerprint-based indoor positioning system," in *Proceedings of the 10th International Conference (ELEKTRO '14)*, pp. 64–67, Rajecke Teplice, Slovakia, May 2014.

[9] Y. Yuan, L. Pei, C. Xu, Q. Liu, and T. Gu, "Efficient WiFi fingerprint training using semi-supervised learning," in *Proceedings of the Ubiquitous Positioning Indoor Navigation and Location Based Service (UPINLBS '14)*, pp. 148–155, Corpus Christ, Tex, USA, November 2014.

[10] M. Youssef and A. Agrawala, "The Horus location determination system," *Wireless Networks*, vol. 14, no. 3, pp. 357–374, 2008.

[11] K. Kaemarungsi and P. Krishnamurthy, "Properties of indoor received signal strength for WLAN location fingerprinting," in *Proceedings of the 1st Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services (MOBIQUITOUS '04)*, pp. 14–23, Boston, MA, USA, August 2004.

[12] L. N. Chen, B. H. Li, K. Zhao, C. Rizos, and Z. Q. Zheng, "An improved algorithm to generate a Wi-Fi fingerprint database for indoor positioning," *Sensors*, vol. 13, no. 8, pp. 11085–11096, 2013.

[13] B. Zhao, L. Pei, C. Xu, and L. Gu, "Graph-based efficient WiFi fingerprint training using un-supervised learning," in *Proceedings of the 28th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS '15)*, pp. 2301–2310, Tampa Convention Center, Tampa, Fla, USA, September 2015.

[14] G. Shakhnarovich, P. Indyk, and T. Darrell, "Nearest-neighbor methods in learning and vision," *IEEE Transactions on Neural Networks*, vol. 19, no. 2, article 377, 2008.

[15] R. M. Faragher and R. K. Harle, "SmartSLAM—an efficient smartphone indoor positioning system exploiting machine learning and opportunistic sensing," *ION GNSS*, vol. 13, pp. 1006–1019, 2013.

[16] Mitchell and M. Tom, "Machine Learning," in *McGraw-Hill Science/Engineering/Math*, 2007, (Bayes Theorem) 156–163, (Naïve Bayes classifier) 177-178, (KNN) 230–234.

[17] L. B. Del Mundo, R. L. D. Ansay, C. A. M. Festin, and R. M. Ocampo, "A comparison of Wireless Fidelity (Wi-Fi) fingerprinting techniques," in *Proceedings of the 2011 International Conference on ICT Convergence (ICTC '11)*, pp. 20–25, Seoul, South Korea, September 2011.

[18] A. H. Salamah, M. Tamazin, M. A. Sharkas, and M. Khedr, "An enhanced WiFi indoor localization System based on machine learning," in *Proceedings of the 2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN '16)*, pp. 1–8, Alcala de Henares, Spain, October 2016.

[19] L. Calderoni, M. Ferrara, A. Franco, and D. Maio, "Indoor localization in a hospital environment using random forest classifiers," *Expert Systems with Applications*, vol. 42, no. 1, pp. 125–134, 2015.

[20] C. Zhou and A. Wieser, "Application of backpropagation neural networks to both stages of fingerprinting based WIPS," in *Proceedings of the 2016 4th International Conference on Ubiquitous Positioning, Indoor Navigation and Location Based Services (UPINLBS '16)*, pp. 207–217, Shanghai, China, November 2016.

[21] M. S. Ifthekhar, N. Saha, and Y. M. Jang, "Neural network based indoor positioning technique in optical camera communication system," in *Proceedings of the 5th International Conference on Indoor Positioning and Indoor Navigation(IPIN '14)*, pp. 431–435, Busan, South Korea, October 2014.

[22] L. Pei, M. Zhang, D. Zou, R. Chen, and Y. Chen, "A survey of crowd sensing opportunistic signals for indoor localization," *Mobile Information Systems*, vol. 2016, Article ID 4041291, 16 pages, 2016.

[23] A. W. Tsui, Y.-H. Chuang, and H.-H. Chu, "Unsupervised learning for solving RSS hardware variance problem in WiFi localization," *Mobile Networks and Applications*, vol. 14, no. 5, pp. 677–691, 2009.

[24] M. B. Kjærgaard and C. V. Munk, "Hyperbolic location fingerprinting: a calibration-free solution for handling differences in signal strength (concise contribution)," in *Proceedings of the 6th Annual IEEE International Conference on Pervasive Computing and Communications (PerCom '08)*, Hong Kong, China, March 2008.

[25] F. Dong, Y. Chen, J. Liu, Q. Ning, and S. Piao, "A calibration-free localization solution for handling signal strength variance," in *Mobile Entity Localization and Tracking in GPS-Less Environments*, pp. 79–90, Springer, Berlin, Germany, 2009.

[26] A. K. M. Mahtab Hossain, Y. Jin, W. Soh, and H. N. Van, "SSD: a robust RF location fingerprint addressing mobile devices' heterogeneity," *IEEE Transactions on Mobile Computing*, vol. 12, no. 1, pp. 65–77, 2013.

[27] J. Correa, E. Katz, P. Collins, and M. Griss, "Room-Level WiFi Location Tracking," MRC-TR-2008-02, Carnegie Mellon Silicon Valley, CyLab Mobility Research Center, 2008.

[28] A. Buchman and C. Lung, "Received signal strength based room level accuracy indoor localisation method," in *Proceedings of the 4th IEEE International Conference on Cognitive Infocommunications (CogInfoCom '13)*, pp. 103–108, Budapest, Hungary, December 2013.

[29] S. N. Patel, K. N. Truong, and G. D. Abowd, "Powerline positioning: a practical sub-room-level indoor location system for domestic use," in *Proceedings of the 8th International Conference (UbiComp '06)*, pp. 441–458, Orange County, CA, USA, 2006.

[30] R. Yasmine and L. Pei, "Indoor fingerprinting algorithm for room level accuracy with dynamic database," in *Proceedings of the 2016 4th International Conference on Ubiquitous Positioning, Indoor Navigation and Location Based Services (UPINLBS '16)*, pp. 113–121, Shanghai, China, November 2016.

[31] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, 1995.

[32] V. N. Vapnik, *Statistical Learning Theory*, Wiley, New York, NY, USA, 1998, ISBN: 978-0-471-03003-4. pp. I–XXIV, 1–736.

[33] T. M. Cover, "Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition," *IEEE Transactions on Electronic Computers*, vol. EC-14, no. 3, pp. 326–334, 1965.

Advances in
Multimedia

The Scientific
World Journal

International Journal of
Distributed
Sensor Networks

Journal of
Industrial Engineering

Applied
Computational
Intelligence and Soft
Computing

Advances in
Fuzzy
Systems

Modelling &
Simulation
in Engineering

Journal of
Computer Networks
and Communications
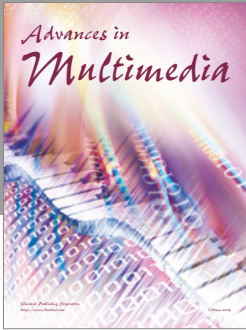
![Hindawi]

Submit your manuscripts at
https://www.hindawi.com

Advances in
Artificial
Intelligence

Advances in
Computer Engineering

International Journal of
Computer Games
Technology

International Journal of
Biomedical Imaging

Advances in
Artificial
Neural Systems

Advances in
Software Engineering

Journal of
Robotics

Advances in
Human-Computer
Interaction

Computational
Intelligence and
Neuroscience

International Journal of
Reconfigurable
Computing

Journal of
Electrical and Computer
Engineering